REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
10-02-2016	Final Report	1-Aug-2007 - 31-Oct-2013

4. TITLE AND SUBTITLE

Final Report: Applying signature extraction and classification algorithms on express on profiles of CD markers and toll like receptors to classify and predict exposures to various pathogens

5a. CONTRACT NUMBER
W911NF-07-1-0479

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
6310AH

6. AUTHORS

Seid Muhie

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES

Georgetown University
37th and O Streets, NW

Washington, DC 20057 -1789

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
53273-LS.4

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

To classify a disease samples using high throughput genomic and proteomic data, it is essential to decide which toll like receptors and CD marker should be included in a predictor list. Too few markers may not be enough to discriminate and classify an exposure. Having too many Markers is not optimal either, as some of these markers may be irrelevant to the diagnosis and may reduce the information decisive factor due to adding noise. Efforts are made to select an optimal set of targets for which to start the training of a set of predictors. This is accomplished by a variety of means such as the neighborhood analysis (Golub et al 1999), principal component analysis (Khan et al

15. SUBJECT TERMS

classification algorithms, pathogens, CD markers

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Rasha Hammamieh
UU	UU	UU	UU		19b. TELEPHONE NUMBER 301-619-2338

Report Title

Final Report: Applying signature extraction and classification algorithms on express on profiles of CD markers and toll like receptors to classify and predict exposures to various pathogens

ABSTRACT

To classify a disease samples using high throughput genomic and proteomic data, it is essential to decide which toll like receptors and CD marker should be included in a predictor list. Too few markers may not be enough to discriminate and classify an exposure. Having too many Markers is not optimal either, as some of these markers may be irrelevant to the diagnosis and may reduce the information decisive factor due to adding noise. Efforts are made to select an optimal set of targets for which to start the training of a set of predictors. This is accomplished by a variety of means such as the neighborhood analysis (Golub et al 1999), principal component analysis (Khan et al 2000), and gene shaving (Hastie et al 2000).

Various algorithms and tools are developed and described in the literature. These algorithms will serve as a foundation for the development of the statistical classification tool.

We will examine these algorithms for best and optimal prediction model and feature extraction.

Initially, data generated using cDNA microarrays will be processed, filtered and analyzed using in house data analysis tools. Expression profiles for the toll like receptors and CD markers for each pathogen at various time points will be extracted. These profiles will be used to identify the markers that are good discriminators for certain pathogen at certain time point. In the process of analyzing the data, we consider two assumptions: 1) The distribution of the gene intensities in a sample is normal and 2) A gene is a good discriminator if it is present at a consistently high level in one class and absent or present at a consistently low level in the other class.

To validate each list of predictors, we will use our database of gene expression as a training set and add some blinded samples to see whether these predictors are able to identify an exposure by analyzing the expression profiles of toll like receptors and CD markers correlated with this exposure.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received Paper

TOTAL:

Number of Manuscripts:

Books

Received Book

TOTAL:

Received Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Seid Muhie	0.20
FTE Equivalent:	**0.20**
Total Number:	**1**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED

FTE Equivalent:
Total Number:

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED

FTE Equivalent:
Total Number:

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:..... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):..... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields: 0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME	PERCENT_SUPPORTED

FTE Equivalent:
Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

Technology Transfer

Applying signature extraction and classification algorithms on express on profiles of CD markers and toll like receptors to classify and predict exposures to various pathogens

- Statement of the problem studied

Pathogen detection and identification tools developed are not always effective especially in early stages post exposure. Host response to biological threat agents has been a very important issue in the case of an outbreak. Identification of signature markers for exposures to various biological threat agents provides a vital tool for classification of outbreaks.

This project attempts to address this need by exploring the feasibility of employing computational methods to determine predictors and classifiers of various pathogens.

We have obtained a large body of experimental data characterizing effects over time for exposure to various biological threat agents. We are establishing a database of gene expression profiles at multiple time points for thousands of genes. In order to diagnose and treat not only known biothreats, but also newly engineered ones, it is important to identify biomolecular unique signatures underlying the observed host response to a pathogen. Computational approaches are essential to organize and visualize the variety of data and to facilitate feature extraction and prediction of an exposure.

- Summary of the most important results

We developed an algorithm to apply predictive modeling and feature extraction using our continuously growing microarray gene expression database obtained by exposing PBMCs to various classes of pathogens (virus, toxin, gram negative and gram positive bacteria) at various time points. We carried out carrying out gene expression analysis for SEB, Dengue, Plague, VEE, Bot toxin, at various time points in more than three replicates each.

Host gene expression in vitro: Microarray analysis was carried out at 3-6 time periods post exposure of PBMC to each pathogen or vehicle. Prior studies [11] showed specific gene sets related to sex, age and other parameters, therefore it was important to first identify genes that are normally variant among healthy humans. Data from only the control samples of these healthy donors were subjected to ANOVA ($p=<0.05$) and 6% of the genes varied widely among the individuals who were healthy human donors. These genes that showed inconsistent expression profiles were excluded from further comparisons among the data sets from both control and exposed samples. This provided a baseline to confidently identify transcriptional responses induced by bacteria (anthrax, plague, *Brucella*), toxins (CT, SEB, BoNTA), or viruses (Dengue, VEE).

Consistency of responses: We used PBMC from at least 3 different donors, exposing cells to pathogen or vehicle for specified periods of time.

Unique gene patterns induced by BTAs: The gene responses were dissected to identify sets of genes that will differentiate one agent from another based on the patterns of host gene induction. The GeneSpring (Silicon Genetics, California) clustering diagram illustrates gene expression patterns that can discriminate among the various pathogenic agents by identification of sets of genes where up regulation and down regulation is seen for specific pathogens. The combination of these selected genes can be the foundation for designing specific diagnostic assays for exposure to one or more agents. Additionally, gene patterns for the earliest exposure for SEB or CT clustered less closely with the later exposure times, but when observed relative to all pathogens, the four exposure time periods for SEB were relatively closely clustered. A striking observation is that for all pathogens except SEB, the longest exposure times differ markedly from the clusters of the early time periods. For *B. anthracis*, *Y. pestis*, *B. melitensis*, and CT, those late exposure times cluster together for these various pathogens. This loss of pathogen-specific responses in vitro after lengthy exposure was not seen for the in vivo studies.

Use of training and test data sets for classifying test exposures: To determine whether the microarray data obtained in this study can be used to predict the exposure type of an uncharacterized sample or condition, we applied a supervised learning method for class prediction (GeneSpring) that uses the k-nearest neighbor algorithm. When algorithm was applied on the data set (training set) to predict the exposure type of a data set obtained from an exposure to *Y. pestis* (test set), we were able to correctly predict the type of exposure with a $p<0.02$. We previously reported that a set of predictor genes was identified when samples from exposures of piglets to SEB were used as test sets [12, 13].

Functional classification of genes differentially regulated:

Gene ontological analysis was carried out for the genes that were differentially expressed. Comparison of gene responses, based on functional similarities, not surprisingly, showed many up regulated genes coding for inflammatory mediators. We clustered and sorted the differentially expressed genes by their functional classification. For gene group (*i*) "Growth Factor, Cytokines & Chemokines," anthrax, *Brucella* and SEB showed major up regulation of most genes coding for inflammatory mediators; the other 5 agents had mixed or modest effects. Similarly, categories (*iii*) "Interleukins and Interferon Receptors" and (*iv*) "Interleukins" showed up regulation by most pathogens, notable exceptions being the viruses. Down regulated genes, though seen extensively throughout the study, displayed functional clustering for each pathogenic agent such as (*ii*) "Homeostasis & detoxification," (*v*) "Ligand-gated ion channels," and so forth. Plague induced high levels

of interleukin-6, macrophage inflammatory protein-1 beta, tumor necrosis factor–alpha (TNF-α), and granulocyte macrophage colony stimulating factor (GM-CSF) when compared with *Brucella* and anthrax. Not surprisingly, the superantigen SEB displayed kinetic patterns for over expression of interferon-γ, IL-2, IL-6, MIP-1α, and GM-CSF. There are major differences in expression of death receptors, homeostasis, and caspases, examples of which include defensins and certain oxidases (homeostasis) that are down regulated by plague and SEB. A large number of transcription factors are down regulated by anthrax, *Brucella*, and SEB, but plague consistently down regulated the widest range of these genes.

Gene responses induced by BTAs in vivo; comparison with in vitro changes: To determine gene changes induced by BTAs in an animal model, NHP were exposed to *B. anthracis* spores by aerosol challenge. This model has been characterized previously to mimic inhalation anthrax in humans. Blood samples were collected 24 h, 48 h, and 72 h post exposure (by 72 h the NHP were beginning to show signs of the illness, which progresses very rapidly to lethality). The gene expression profiles for in vitro exposure of PBMC to anthrax spores were compared with those found in isolated PBMC at various time periods from NHP. Even by 24 h, a robust response was observed, showing up regulation of genes coding for proteases; proteosome components c2, c3, c5; various cytokines; pro-apoptotic genes; cyclic adenosine monophosphate (cAMP)-related kinases, cAMP regulated transcription factors; and hypoxia inducible factor-1 (HIF-1). Down regulated genes included tyrosine kinases, cytokine receptors, growth factors, and adenosine diphosphate (ADP) ribosylation factors. Comparison of the in vivo results with the in vitro changes induced by anthrax, showed remarkable similarities in gene patterns. Clearly many more changes were observed in vivo than in vitro. Certain surface antigens showed significant alteration that was unique to anthrax exposure. Diagrams were constructed to identify sets of genes that were up regulated at either 24 or 72 h; other gene sets showed up regulation at both time periods.

A few genes were selected that showed changes induced by *B. anthracis* exposure were confirmed by RT-PCR, and the level of expression was compared both in vitro and in vivo after anthrax exposure. Altered regulation of that G-protein was not seen with the other pathogenic agents. In an experiment of SEB exposure to NHP, IL-6 and guanylate binding protein GBP-2 were up regulated (6- and 65-fold, respectively) by 30 min post-exposure and the increased expression persisted through 24 h .

- Bibliography

1. Gibb TR, Norwood DA, Jr., Woollen N, Henchal EA: Viral replication and host gene expression in alveolar macrophages infected with Ebola virus (Zaire strain). Clin Diagn Lab Immunol 2002, 9:19-27.
2. Henchal EA, Teska JD, Ludwig GV, Shoemaker DR, Ezzell JW: Current laboratory methods for biological threat agent identification. Clin Lab Med 2001, 21:661-678.
3. Hurtle W, Lindler L, Fan W, Shoemaker D, Henchal E, Norwood D: Detection and identification of ciprofloxacin-resistant Yersinia pestis by denaturing high-performance liquid chromatography. J Clin Microbiol 2003, 41:3273-3283.
4. http://www.cdc.gov/ncidod/dbmd/diseaseinfo.htm
5. Guarner J, Shieh WJ, Greer PW, Gabastou JM, Chu M, Hayes E, Nolte KB, Zaki SR: Immunohistochemical detection of Yersinia pestis in formalin-fixed, paraffin-embedded tissue. Am J Clin Pathol 2002, 117:205-209.
6. Sun P, Celluzzi CM, Marovich M, Subramanian H, Eller M, Widjaja S, Palmer D, Porter K, Sun W, Burgess T: CD40 ligand enhances dengue viral infection of dendritic cells: a possible mechanism for T cell-mediated immunopathology. J Immunol 2006, 177:6497-6503.
7. Das R, Dhokalia A, Huang XZ, Hammamieh R, Chakraborty N, Lindler LE, Jett M: Study of proinflammatory responses induced by Yersinia pestis in human monocytes using cDNA arrays. Genes Immun 2007, 8:308-319.
8. Glantz SA, Slinker BK: Primer of Applied Regression and Analysis of Variance. New York: McGraw-Hill; 1990.
9. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP: Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Research 2002, 30:e15-e15.
10. Eisen MB, Spellman PT, Brown PO, Botstein D: Cluster analysis and display of genome wide expression patterns. Proc Natl Acad Sci U S A 1998, 95:14863-14868.
11. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO: Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci U S A 2003, 100:1896-1901.
12. Hammamieh R, Bi S, Das R, Neill R, Jett M: Modeling of SEB-induced host gene expression to correlate in vitro to in vivo responses. Biosens Bioelectron 2004, 20:719-727.
13. van Gessel YA, Mani S, Bi S, Hammamieh R, Shupp JW, Das R, Coleman GD, Jett M: Functional piglet model for the clinical syndrome and postmortem findings induced by staphylococcal enterotoxin B. Exp Biol Med (Maywood) 2004, 229:1061-1071.
14. Franz DR, Jahrling PB, Friedlander AM, McClain DJ, Hoover DL, Bryne WR, Pavlin JA, Christopher GW, Eitzen EM, Jr.: Clinical recognition and management of patients exposed to biological warfare agents. Jama 1997, 278:399-411.
15. Bayani J, Brenton JD, Macgregor PF, Beheshti B, Albert M, Nallainathan D, Karaskova J, Rosen B, Murphy J, Laframboise S, et al: Parallel analysis of sporadic primary ovarian carcinomas by spectral karyotyping, comparative genomic hybridization, and expression microarrays. Cancer Res 2002, 62:3466-3476.
16. Li S, Ross DT, Kadin ME, Brown PO, Wasik MA: Comparative genome-scale analysis of gene expression profiles in T cell lymphoma cells during malignant progression using a complementary DNA microarray. Am J Pathol 2001, 158:1231-1237.
17. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, et al: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001, 98:10869-10874.

18. Boldrick JC, Alizadeh AA, Diehn M, Dudoit S, Liu CL, Belcher CE, Botstein D, Staudt LM, Brown PO, Relman DA: Stereotyped and specific gene expression programs in human innate immune responses to bacteria. Proc Natl Acad Sci U S A 2002, 99:972-977.
19. Huang Q, Liu D, Majewski P, Schulte LC, Korn JM, Young RA, Lander ES, Hacohen N: The plasticity of dendritic cell responses to pathogens and their components. Science 2001, 294:870-875.
20. Friedlander AM: Anthrax: clinical features, pathogenesis, and potential biological warfare threat. Curr Clin Top Infect 2000, 20:335-349.
21. MacDonald GH, Johnston RE: Role of dendritic cell targeting in Venezuelan equine encephalitis virus pathogenesis. J Virol 2000, 74:914 - 922.
22. Campbell GL, Dennis DT: Plague and other Yersinia infections. In Harrison's principles of internal medicine. Volume 14th ed. Edited by Kasper DL ea. New York: McGraw Hill; 1998: 975-983